

INTERSEZIONI

Digital humanities e organizzazione della conoscenza: una pratica di insegnamento nel LODLAM

di Francesca Tomasi

Introduzione

Fin dalle origini della cosiddetta *humanities computing*, altrimenti declinata in Italia con il termine ‘informatica umanistica’, numerosi sono stati i contributi critici che hanno tentato di fare il punto su ciò che significa insegnare questa disciplina (se così possiamo chiamarla). Proposte diverse sono emerse negli anni, che spaziano da soluzioni pratiche – a titolo di esempio, potremmo menzionare il progetto Aco*Hum¹ o anche la proposta tedesca di un curriculum di settore² –, al ragionamento teorico sulla dimensione pedagogica e formativa di questi saperi trasversali³.

Lavorare nel settore di quelle che, più di recente, sono state chiamate *digital humanities* (da ora DH)⁴ significa aggregare conoscenze eterogenee, provenienti da numerose discipline tradizionali, con l’obiettivo di elaborare nuove strategie di rappre-

FRANCESCA TOMASI, Alma mater studiorum Università di Bologna, Dipartimento di filologia classica e italianistica, e-mail francesca.tomasi@unibo.it.

Ultima consultazione siti web: 26 maggio 2020.

1 *Computing in humanities education: a European perspective*, edited by Koenraad De Smedt [et al.]. Bergen: University of Bergen. HIT-centre, 1999, <<http://korpus.uib.no/humfak/AcoHum/book>>.

2 Manfred Thaller; Patrick Sahle, *Digitale Geisteswissenschaften*. Cologne: Cologne Center for eHumanities, 2011, <<http://www.cceh.uni-koeln.de/Dokumente/BroschuereWeb.pdf>>.

3 Si vedano, a titolo di esempio: Stephen Brier, *Where’s the pedagogy? The role of teaching and learning in the digital humanities*. In: *Debates in the digital humanities*, new edition, edited by Matthew K. Gold. Minneapolis: University of Minnesota Press, 2012, p. 390-412, <<https://www.jstor.org/stable/10.5749/j.ctttv8hq.25>>; Geoffrey Rockwell; Stefan Sinclair, *Acculturation and the digital humanities community*. In: *Digital humanities pedagogy: practices, principles, politics*, edited by Brett D. Hirsch. Cambridge: Open Book Publishers, 2012, p. 177-211, <<https://www.openbookpublishers.com/product/161>>; *Teaching with digital humanities: tools and methods for Nineteenth-Century American literature*, edited by Jennifer Travis, Jessica DeSpain. Urbana: University of Illinois Press, 2018, <<https://www.jstor.org/stable/10.5406/j.ctv8bt13m>>.

4 La prima occorrenza del termine *digital humanities* è il titolo del *companion* edito da Blackwell nel 2004 (*A companion to digital humanities*, edited by Susan Schreibman, Ray Siemens, John Unsworth. Oxford: Blackwell, 2004, <<http://www.digitalhumanities.org/companion>>).

sentazione, analisi e disseminazione del sapere; ma anche con lo scopo di utilizzare strumenti, tecniche e tecnologie per fare ricerca con una nuova visione delle discipline umanistiche: la coscienza della computabilità del dato umanistico.

Insegnare DH significa fornire quelle competenze e capacità necessarie per convertire queste conoscenze in pratica, tanto sul piano metodologico quanto su quello tecnico-applicativo, con l'obiettivo di studiare le discipline umanistiche attraverso nuove lenti di osservazione, finalizzate ad agevolare la produzione di risultati innovativi e non convenzionali.

Nel 2017 l'Università di Bologna ha avviato un corso di laurea magistrale internazionale, erogato integralmente in lingua inglese, in Digital humanities and digital knowledge (da ora DHDK)⁵. Il corso di laurea è di natura altamente interdisciplinare, ed è stato ideato nell'ottica di una profonda e mutua interazione tra discipline umanistiche e informatica, aggiungendo il contributo teorico proveniente da alcune discipline trasversali come diritto, economia, comunicazione e strategia sociale. Il corso di laurea DHDK focalizza in particolare sull'esigenza di applicare conoscenze e competenze all'interno e per un progetto; l'obiettivo principale è infatti quello di trasmettere agli studenti il principio di fare ricerca applicando metodi e modelli DH per realizzare oggetti, ambienti e strumenti digitali scientificamente solidi e in modo orientato a risolvere reali problemi di ricerca⁶. Ibridare conoscenze per creare una nuova coscienza nello studio delle *humanities* è l'obiettivo della formazione erogata durante il corso di studio. L'apprendimento è di conseguenza inteso come la capacità di ragionare criticamente su tali conoscenze per produrre risultati di ricerca consistenti e originali⁷.

Durante il primo anno di corso, gli studenti iniziano ad avvicinarsi al tema DH attraverso alcuni saperi trasversali: la filologia digitale; le scienze del libro e del documento; i sistemi di analisi, recupero ed estrazione di stringhe di caratteri; i musei e la museologia; i testi digitali nelle discipline umanistiche e infine i sistemi di organizzazione della conoscenza. Oltre a queste discipline, viene fornito un *know-how* informatico più specifico: il pensiero computazionale, i linguaggi di programmazione, i database multimediali, le tecnologie web, la rappresentazione ed estrazione della conoscenza, in particolare nell'ottica del web semantico; i metodi multimediali e tridimensionali nel contesto della valorizzazione dei beni culturali materiali.

La maggior parte di queste attività ha elaborato un metodo di valutazione basato sulla realizzazione di un progetto, che, come si diceva, deve partire da uno scenario reale, ovvero da un autentico quesito di ricerca. Dato un tema, un argomento, un obiettivo o uno scopo, agli studenti – preferibilmente organizzati in gruppo – viene chiesto di progettare un lavoro orientato a produrre risultati scientificamente solidi e che possano essere poi presentati, o disseminati, attraverso supporti eterogenei (ad esempio un articolo scientifico, un'applicazione web, un ambiente digitale o una mostra virtuale).

⁵ Sito ufficiale di DHDK, <<https://corsi.unibo.it/2cycle/DigitalHumanitiesKnowledge>>. Una lista completa dell'offerta formativa nelle DH si può leggere sul DARIAH registry, <<https://dhcr.clarin-dariah.eu>>.

⁶ Un recente progetto, DHKey 2020, vuole essere una piattaforma (che a sua volta è basata su un dataset RDF) per raccogliere i progetti realizzati dagli studenti di DHDK come prova d'esame dei diversi insegnamenti della LM. La piattaforma è stata realizzata da un laureato di DHDK, Fabio Mariani, e implementata grazie a un *grant* erogato dal centro di ricerca bolognese sulle DH chiamato DH.arc, <<https://projects.dharc.unibo.it/dhdkey>>. Il dataset verrà popolato nel corso del 2020.

⁷ Francesca Tomasi, *Digital humanities and digital knowledge (DHDK): international second cycle/master degree*, «Umanistica digitale», 2 (2018), p. 225-231, <<https://umanisticadigitale.unibo.it/article/view/7862>>.

Ogni progetto documenta l'ibridazione delle conoscenze apprese studiando in modo trasversale le discipline erogate durante il primo anno di formazione dentro DHDK. Filologia, letteratura, linguistica, scienza del libro e del documento, patrimonio materiale insieme al pensiero computazionale, la logica di programmazione, le tecnologie web e la gestione della conoscenza, diventano dunque i punti di partenza per avviare un progetto nel settore delle DH.

Lo schema virtuale di valutazione per ciascun progetto segue questa griglia ideale: - originalità e consistenza dell'idea alla base del progetto (scopo, contesto, profilo dell'utente finale);

- qualità e correttezza della metodologia adottata (strumenti, metodi, modelli);

- capacità di proporre e gestire un lavoro di squadra (funzioni dei membri del team di lavoro);

- valore dell'output finale (caratteristiche, tecnologie, servizi).

Il corso di Knowledge organization and cultural heritage (che per brevità chiameremo KO and CH) è concepito come un'opportunità per combinare abilità e competenze eterogenee, provenienti da diverse discipline nel dominio DH, necessarie a comprendere come valorizzare i dati culturali provenienti da archivi, biblioteche e musei per fornire conoscenze reali, e quindi spendibili, all'utente finale (che sia utente umano o macchina).

Insieme alla consapevolezza della *knowledge organization* (da ora KO) come metodologia⁸, questo corso si concentra sul dominio *libraries, archives and museums* (da ora LAM), scegliendo i linked open data (da ora LOD) come la modalità per organizzare i dati prodotti dalle istituzioni culturali, e con lo scopo didattico finale di superare barriere e steccati disciplinari, ovvero con il fine di andare «beyond the silos of the LAMs»⁹, al contempo utilizzando le teorie, le metodologie e le tecniche del DH¹⁰.

L'idea è di fornire agli studenti gli strumenti necessari per comprendere come passare dai tradizionali *finding aid* archivistici o dai cataloghi bibliografici, ad ambienti a base semantica esplicita. Un grafo della conoscenza nel contesto del patrimonio culturale (o anche un *cultural heritage knowledge graph*) è l'obiettivo finale di questo ragionamento critico su descrizione, gestione, diffusione, uso e riuso dei dati, in una più generale visione della manipolazione delle risorse all'interno del complesso ciclo vitale degli oggetti informativi.

Un approccio orientato alla focalizzazione su dati, metadati e, in generale, sugli standard di contenuto e struttura nel dominio LAM, consente agli studenti di acquisire una nuova visione dei problemi legati alla descrizione del patrimonio culturale.

Attraverso la nozione di *semantic web* poi, in particolare nel contesto della modellizzazione concettuale, gli studenti acquisiscono la capacità di riflettere sul patrimonio culturale, attribuendo un nuovo significato al principio della valorizzazione.

8 Birger Hjørland, *What is knowledge organization (KO)?*, «Knowledge organization», 35 (2008), n. 2/3, p. 86-101, <<https://www.nomos-elibrary.de/10.5771/0943-7444-2008-2-3-86/what-is-knowledge-organization-ko-volume-35-2008-issue-2-3?page=1>>.

9 Diane M. Zorich; Gunter Waibel; Ricky Erway, *Beyond the silos of the LAMs: collaboration among libraries, archives and museums*. Dublin (OH): OCLC Research, September 2008, <<http://www.oclc.org/content/dam/research/publications/library/2008/2008-05.pdf>>.

10 Sull'interesse delle DH nei confronti di LOD e biblioteche, si veda ad esempio l'attività di ADHO come espressa nei gruppi SIG (*special interest group*) su queste tematiche, <<https://adho.org/sigs>>.

Con le DH si aggiunge poi la consapevolezza del senso dell'ermeneutica attraverso la pratica della produzione di nuova conoscenza come atto interpretativo in una dimensione progettuale.

Per concludere, lavorare su scenari reali è la chiave per raggiungere gli obiettivi educativi enunciati.

Scendendo nel dettaglio dell'offerta formativa, l'idea di valutare l'apprendimento nel settore KO and CH può essere riassunta come segue: il *workflow* inizia con l'osservazione di una porzione di dominio LAM, che nasca dall'idea di un possibile progetto (selezione dei dati); il *requirement* è di procedere per astrazione, disegnando un primo modello, capace di descrivere quanto osservato; il passo successivo è l'identificazione dei metadati e degli standard utilizzati dalle istituzioni da cui i dati sono acquisiti; l'osservazione deve quindi essere arricchita con nuovi fenomeni descrittivi, esito di un procedimento di analisi; successivamente, il processo prevede la traduzione del primo modello in una rappresentazione concettuale, riutilizzando ontologie pertinenti; infine, per completare il flusso di lavoro, è necessario produrre LOD in una delle possibili serializzazioni RDF e aggiungere collegamenti ai dati già presenti nel *LOD cloud*.

Detto altrimenti, dato un dominio, rappresentato da un insieme di oggetti analogici (ma già acquisiti digitalmente e descritti secondo gli standard in essere), provenienti dal dominio bibliografico, archivistico e museale/materiale, gli studenti sono tenuti ad arricchire le descrizioni esistenti riflettendo sulle possibili relazioni trasversali tra le entità coinvolte nell'analisi. Vengono per questo utilizzati strumenti propri del web semantico e metodi canonici dell'approccio LOD, ma gli studenti sono anche invitati a formalizzare la propria interpretazione sul dominio osservato, arricchendolo con nuovi collegamenti semantici, anche a risorse esterne.

Al fine di riflettere su questi aspetti, il presente articolo è organizzato come segue: vengono presentati i contenuti del corso KO and CH; viene descritta una proposta di linee guida per un progetto LODLAM; per concludere viene simulato uno scenario reale, ovvero una possibile idea progettuale.

Knowledge organization and cultural heritage: il programma del corso

Il corso KO and CH ha lo scopo di introdurre alle problematiche dell'organizzazione della conoscenza. In particolare, si suppone che alla fine del corso gli studenti siano in grado di: utilizzare gli standard del dominio LAM per la descrizione dei dati culturali; conoscere le pratiche in uso per l'arricchimento degli stessi dati attraverso strategie semantiche (ontologie e vocabolari controllati); creare metadati complessi per oggetti digitali nel dominio del patrimonio culturale.

Il corso si concentra su tre argomenti strettamente correlati:

- lo studio dei principali standard internazionali nel settore LAM;
- la riflessione sul concetto di organizzazione della conoscenza (KO) e sui sistemi di organizzazione della conoscenza (KOS), e in particolare sul ruolo delle ontologie come modelli concettuali;
- l'analisi delle attività semantiche attraverso metodi KO nel contesto del LAM, con particolare attenzione al ruolo del movimento LOD.

In una frase, il corso mira a fornire agli studenti le competenze per creare, gestire, diffondere e preservare metadati complessi in LOD, come metodo KO nel dominio LAM.

Per offrire una panoramica delle pratiche e dei metodi in uso nel LAM a livello nazionale e internazionale, durante il corso vengono discusse le attività delle tre organizzazioni di settore, e in particolare vengono presentati gli standard elaborati.

Quindi, per il dominio biblioteche, viene esaminato il ruolo dell'IFLA¹¹ e l'attività della Library of Congress¹², insieme agli standard italiani proposti da ICCU¹³; per gli archivi, è l'attività dell'ICA¹⁴, insieme all'ICAR¹⁵, l'oggetto di interesse; per i musei viene analizzata la proposta dell'ICOM¹⁶, insieme alla gestione della descrizione del patrimonio culturale come disegnata dall'ICCD¹⁷.

Partendo dunque dal ruolo delle istituzioni culturali nell'organizzazione dell'informazione, come la funzione ancestrale dei luoghi di conservazione del patrimonio¹⁸, il mondo di LAM è allora presentato come un grafo globale di dati, meta-dati e vocabolari¹⁹.

L'analisi degli standard è particolarmente importante per affrontare una serie di problemi, vale a dire per acquisire consapevolezza del fatto che:

- ogni istituzione culturale ha sviluppato i propri standard;
- ogni istituzione ha alcune tipologie di oggetti, o anche fonti materiali, da preservare e descrivere;
- ogni oggetto, o anche ogni tipologia documentaria, ha il suo standard da seguire;
- non tutti gli oggetti (dai libri, antichi e moderni, ai documenti archivistici nelle loro varie tipologie; dai sigilli alle mappe; dalle riviste alle statue; dai manoscritti ai dipinti) hanno uno standard di riferimento per la descrizione;
- alcuni standard sono condivisi da diverse istituzioni.

Lo sforzo di classificare gli standard per i metadati²⁰ è un passo importante per affrontare questi problemi. Tuttavia, nella dimensione del web semantico, questa classificazione è sfuggente, in particolare per quanto riguarda la differenza tra dati e metadati nella produzione di LOD, e di questo è necessario acquisire la giusta consapevolezza.

11 International Federation of Library Associations and Institutions, *IFLA standards*. 9th July 2015, <<https://www.ifla.org/standards>>.

12 Library of Congress, *Authorities*. 25th September 2019, <<https://authorities.loc.gov>>.

13 Istituto centrale per il catalogo unico, *Normative catalografiche, linee guida e standard*. <<https://www.iccu.sbn.it/it/normative-standard>>.

14 International Council on Archives, *International standards*. <<https://www.ica.org/en/public-resources/standards>>.

15 Istituto centrale per gli archivi, *Standard internazionali*. 14 luglio 2020, <<http://www.icar.beniculturali.it/index.php?id=52>>.

16 International Council of Museums, *Standards*. <<https://icom.museum/en/resources/standards-guidelines/standards>>.

17 Istituto centrale per il catalogo e la documentazione, *Standard catalografici*. <<http://www.iccd.beniculturali.it/it/standard-catalografici>>.

18 Arlene G. Taylor, *The organization of information*. Englewood (CO): Libraries Unlimited, 1999.

19 Eero Hyvönen, *Publishing and using cultural heritage linked data on the semantic web*. [San Rafael (CA)]: Morgan & Claypool, 2012, <<https://www.morganclaypool.com/doi/abs/10.2200/S00452ED1V01Y201210WBE003>>.

20 Jenn Riley, *Seeing standards: a visualization of the metadata universe*. 2009-2010, <<http://jennriley.com/metadata-map/>>.

Per ragionare quindi in ottica LOD dal punto di vista teorico, il modello RDF viene introdotto nei termini di una nuova modalità di espressione dei metadati in dimensione tanto sintattica (la tripla), quanto semantica (i predicati ontologici), ragionando sulle principali entità oggetto della rappresentazione di dati aperti e interconnessi: le persone, i luoghi, le date, i concetti.

Quindi, al fine di comprendere i processi che consentono la trasformazione di cataloghi, inventari o ausili di ricerca in generale, in dati che non sono solo aperti, ma interconnessi attraverso tecnologie semantiche o anche link 'tipizzati', vengono presentati e discussi alcuni casi di studio.

L'idea che sta alla base del percorso di apprendimento è quella di lavorare sull'intero flusso di lavoro orientato alla creazione di LOD, per gestire i diversi livelli del processo. La scelta di un vero progetto scientifico, ovvero Zeri & Lode²¹, è lo strumento immaginato tanto per ragionare sulle diverse tipologie di metadati coinvolti in un'attività orientata alla descrizione di risorse culturali eterogenee (foto e opere d'arte, ma anche lettere, cataloghi d'asta e fondi bibliografici), quanto per simulare un *workflow* che mostri come da un database relazionale si passi alla realizzazione di un dataset semantico.

Ma prima di arrivare all'analisi di Zeri & Lode, per dare un'idea molto generale di questo approccio ai LOD, il primo caso di studio discusso durante il corso è Europeana.eu. Questo progetto consente di riflettere sul ruolo del Dublin core (DC) come standard comune, ma anche come vocabolario agnostico e trasversale. Il DC fa emergere anche l'importanza di lavorare non solo su una 'manifestazione' dell'"espressione" di un oggetto culturale, ma su diversi livelli. Attraverso l'approccio FRBR si prende coscienza del fatto che ogni livello richiede i propri standard, strettamente correlati alla tipologia documentaria e alle sue diverse forme comunicative (l'idea di opera, il lavoro sul contenuto, la descrizione di una specifica versione materiale e l'oggetto analogico). Il progetto Europeana LOD²² è utile poi per iniziare a comprendere il principio della produzione di nuovi tipi di dati che ragionino per triple o anche per asserti semantici, e il modello di dati Europeana, l'Europeana data model (EDM)²³, consente di avvicinare alla nozione di ontologia, e di integrazione di modelli ontologici diversi, nella prospettiva LOD²⁴.

Collezioni di vocabolari (ad esempio LOV), *repository* di LOD (ad esempio Datahub), set di dati (ad esempio il *LOD cloud*) e progetti nel dominio LODLAM (ad esempio lod-

21 Ciro Mattia Gonano [et al.], *Zeri e LODE: extracting the Zeri photo archive to linked open data: formalizing the conceptual model*. In: 2014 IEEE/ACM Joint Conference on Digital Libraries (London, United Kingdom, 8th-12th September 2014). [Piscataway]: Institute of Electrical and Electronics Engineers, 2014, p. 289-298, <<https://ieeexplore.ieee.org/document/6970182>>. Marilena Daquino [et al.], *Enhancing semantic expressivity in the cultural heritage domain: exposing the Zeri Photo Archive as linked open data*, «Journal on computing and cultural heritage», 10 (2017), n. 4, article 21, <<https://dl.acm.org/doi/10.1145/3051487>>.

22 Europeana ha realizzato il suo *LOD pilot*, consultabile all'indirizzo <<https://data.europeana.eu>>.

23 Europeana data model (EDM), <<http://pro.europeana.eu/page/edm-documentation>>.

24 Punto di partenza è il volume: Tom Heath; Christian Bizer, *Linked data: evolving the web into a global data space*. [San Rafael (CA)]: Morgan & Claypool, 2011, <<https://www.morganclaypool.com/doi/abs/10.2200/S00334ED1V01Y201102WBE001>>.

lam.org)²⁵, sono riferimenti per ragionare su interscambio semantico, arricchimento e riuso in un mondo fatto di «datasets, value vocabularies e metadata element sets»²⁶.

Un paio di ultime riflessioni sulle modalità di erogazione della didattica e di valutazione del processo di apprendimento.

Il corso si articola in lezioni tradizionali – pensate per garantire l’acquisizione di competenze sugli aspetti teorici della disciplina –, ed esercitazioni di laboratorio – necessarie ad avvicinarsi alla parte pratica del processo –, a cui si affiancano riflessioni sui principi cardine della progettazione e sulle problematiche del lavoro in gruppo.

L’esame finale prevede la discussione di una risorsa esistente nel contesto LODLAM, condotta analizzando caratteristiche e implicazioni teoriche, metodologiche e tecnologiche²⁷ e dalla presentazione di un progetto originale relativo a KO, LOD e LAM. Agli studenti viene chiesto di organizzarsi in gruppi per ideare e realizzare il proprio progetto. Il contributo personale di ciascun membro del team viene valutato durante il colloquio orale, quando viene presentato il lavoro.

La sezione che segue vuole quindi presentare le linee guida per la realizzazione di un progetto nel contesto LODLAM, così come introdotto, e descritto, agli studenti del corso.

Linee guida per un progetto LODLAM

Agli studenti è richiesto di presentare un progetto realizzando un report (preferibilmente una pagina web), in cui dimostrare la capacità di lavorare su un processo tripartito: elaborare un’idea progettuale, scegliere una porzione di dominio e analizzarla (analisi del dominio); elaborare i modelli necessari a descrivere quanto osservato (organizzazione della conoscenza: elaborare i modelli) e quindi produrre dati semantici (rappresentazione della conoscenza: creazione dei dati).

Queste linee guida sono state ideate nei termini di un *workflow* che, partendo da dati reali, proceda all’astrazione, traduca l’interpretazione della realtà osservata nella concettualizzazione attraverso un modello teorico e che dalla modellazione, come procedimento astrattivo, torni sui dati per la verifica dell’applicabilità del modello stesso, arricchito dall’interpretazione.

Analisi del dominio

- Idea. Il requisito per questa prima fase del lavoro è di enucleare un’idea, che possa ad esempio essere il soggetto di una pagina di Wikipedia – come un personaggio storico, un luogo significativo per una qualche ragione, un concetto, un evento importante o anche una data speciale. L’idea deve essere collegata a differenti tipologie di oggetti del dominio LAM – ad esempio documenti d’archivio, libri, artefatti cultu-

²⁵ Linked open vocabularies (LOV), <<http://lov.okfn.org/dataset/lov>>; DataHub, <<https://datahub.io/dataset>>; Linking open data cloud diagram, <<http://lod-cloud.net>>; LODLAM, <<http://lodlam.net>>.

²⁶ Antoine Isaac [et al.], *Library linked data incubator group: datasets, value vocabularies, and metadata element sets: W3C Incubator Group report*. October 2011, <<https://www.w3.org/2005/Incubator/llld/XGR-llld-vocabdataset-20111025>>.

²⁷ Alcune riviste di settore sono suggerite agli studenti al fine di permettere loro di consultare altri progetti potenzialmente affini al proprio ambito di interesse, riuscendo quindi ad arricchire l’esposizione del progetto finale, per esempio: *Journal of information science*, <<http://journals.sagepub.com/home/jis>>; *Knowledge organization*, <<http://www.isko.org/ko.html>>; *Library and information science research*, <<http://www.sciencedirect.com/science/journal/07408188>>; *Journal of documentation*, <<https://www.emerald.com/insight/publication/issn/0022-0418>>.

rali –, ovvero oggetti capaci, a vario livello e sulla base di diversi tipi di relazioni, di esprimere l'idea. Il suggerimento è di raccogliere tutti gli elementi necessari dalla pagina Wikipedia (dati utili, link esterni, riferimenti bibliografici, riferimenti correlati, controllo di autorità laddove presente) al fine di comprendere come arricchire poi i propri dati con altre risorse esistenti. Se l'idea è troppo estesa, si suggerisce di ridurre il dominio, individuando alcuni concetti chiave dello scenario.

- Scelta degli *item*. Dopo aver descritto il proprio scenario, si deve passare alla ricerca di almeno dieci differenti *item* collegati all'idea, cercando di mescolare tipologie documentarie diverse: libri, documenti d'archivio, oggetti materiali (fotografie, mappe, sculture, edifici, manoscritti, sigilli ecc.) in possibili diverse manifestazioni (immagini, audio, video, testi o anche solo metadati descrittivi). Alcuni dettagli utili:

- gli oggetti devono già essere descritti sul web;
- le descrizioni devono provenire dai repertori delle istituzioni di conservazione (ad esempio un libro descritto in OPAC SBN o WorldCat, un manoscritto reperito su Manus, un documento archivistico sul SAN, una fotografia su SIGECweb e più in generale si può ricorrere ai principali aggregatori come Europeana, WDL, CulturalItalia, Internet culturale ecc.)²⁸;
- le descrizioni devono includere informazioni relative a diverse entità potenzialmente coinvolte: persone, luoghi, dati e soggetti o concetti.

Organizzazione della conoscenza: l'elaborazione dei modelli

- Mappa concettuale e modello E/R. Il primo step della modellazione è l'elaborazione di una mappa concettuale, capace di spiegare la propria idea, o anche di rappresentare il proprio scenario. La mappa deve rappresentare i dati reali, così come acquisiti dalle descrizioni degli *item* individuati, assieme alle possibili relazioni. Il secondo passo è la traduzione della mappa in un modello E/R (entità, relazioni e attributi) o realizzata usando un altro formalismo (sempre a livello di grafo). Si passa quindi dalla rappresentazione dei dati reali della mappa, alle entità astratte che configurano lo scenario (persone, luoghi, date e soggetti/concetti) e le relazioni fra di esse.

- Analisi degli standard di metadati e allineamento. L'analisi dei metadati deve essere condotta cercando di identificare gli standard, a livello di contenuto e struttura, utilizzati dalle istituzioni che conservano gli *item*, ed enucleare i descrittori. Il secondo step è l'allineamento dei modelli. In particolare, si richiede di mappare quei descrittori che riportino informazioni relative a persone, luoghi, dati e soggetti/concetti. DC e DCTerms possono essere usati come guida per definire il set delle categorie per l'allineamento. Gli elementi mancanti da DC possono essere aggiunti quando gli standard in uso dalle istituzioni siano più ricchi ed espressivi.

- Modello teorico. Un passo chiave del processo di organizzazione della conoscenza è quello legato all'interpretazione dei dati, atto inteso come esito dell'osservazione degli *item* e delle loro relazioni latenti. L'obiettivo di questo step è di produrre un modello teo-

²⁸ OPAC SBN, <<https://opac.sbn.it>>; Worldcat, <<https://www.worldcat.org>>; Manus, <<https://manus.iccu.sbn.it>>; Sistema archivistico nazionale (SAN), <<https://san.beniculturali.it>>; SIGECweb, <<http://www.catalogo.beniculturali.it>>; Europeana, <<https://www.europeana.eu>>; World digital library (WDL), <<https://www.wdl.org/en>>; CulturalItalia, <<http://www.culturalitalia.it>>; Internet culturale, <<http://www.internetculturale.it>>.

rico, espresso in linguaggio naturale o in pseudo-codice, da cui emergano altri elementi utili a qualificare le proprietà degli *item* selezionati, proprietà che vadano oltre le descrizioni così come acquisite dagli istituti di conservazione. È possibile riutilizzare le informazioni provenienti dalla voce originaria di Wikipedia e trarre ispirazione, per altri descrittori o predicati, da Wikidata e Dbpedia, ma si consiglia di cercare sul web altre fonti per l'arricchimento dei dati e per la riflessione sulle possibili relazioni fra dati.

In particolare, il modello deve essere in grado di rispondere a determinate questioni, che recuperano la classificazione in persone, luoghi, date e soggetti/concetti come entità base dello scenario:

- Chi (persone). Ovvero, quali informazioni possono essere utili per arricchire la descrizione delle persone. Per esempio, quale ruolo assume una persona nel ciclo vitale dell'oggetto culturale scelto fra i propri *item*? Quali dati prosopografici possono essere utili per descrivere una persona? Che tipo di relazioni quella persona può avere rispetto ad altre persone o ad altri *item* del proprio scenario?
- Dove (luoghi). Ovvero, quali informazioni possono essere utili per arricchire la descrizione dei luoghi. Per esempio, può essere utile una geolocalizzazione degli elementi spaziali? Quali possibili declinazioni del concetto di luogo possono essere previste? È il luogo un elemento condiviso da più *item*, magari con accezioni diverse?
- Quando (date). Ovvero, quali informazioni possono essere utili per arricchire la descrizione del concetto di data. Quanti tipi di date diverse ci sono in un *item*? Possono esserci altre date utili a descrivere gli *item*? È la data un elemento chiave per aggregare *item* diversi?
- Cosa (soggetti/concetti). Ovvero, quali informazioni possono essere utili per arricchire la descrizione dell'*item* a livello di contenuto trasmesso? Quali sono le parole chiave che identificano l'*item*, ovvero quali sono i soggetti in grado di descriverlo? C'è un evento in particolare che può essere associato agli *item* in termini di concetto correlato? Posso aggiungere un codice di classificazione per descrivere il contenuto di un *item*?

A questo punto è possibile tornare a un livello di astrazione e quindi affinare il modello E/R e creare un nuovo grafo, che sia il risultato dell'interpretazione espressa nel modello teorico, aggiungendo nuove entità, relazioni e attributi, capaci di descrivere l'osservazione arricchita del dominio.

- Modello concettuale. Fatte queste azioni, si può passare dal modello teorico a una rappresentazione formale, ovvero alla concettualizzazione necessaria a stabilire classi e proprietà di una propria ontologia. Il modello concettuale deve riutilizzare schemi, vocabolari e ontologie già esistenti (RDF/RDFS/OWL/SKOS; DC, DCTerms, FRBR, EDM, RDA, OAD, EAC-CPF)²⁹. Il linguaggio naturale o lo pseudo-codice del modello teorico deve essere quindi ripensato in termini di classi e proprietà.

Allo scopo di esprimere l'azione di concettualizzazione si suggerisce di nuovo l'uso di un sistema di formalizzazione a grafo (ad esempio ricorrendo a Grafoo, il «graphical framework for OWL ontologies»)³⁰.

29 Per un elenco di ontologie nel dominio LAM, si veda Marilena Daquino, *Mining authoritativeness in art historical photo archives: semantic web applications for connoisseurship*. Amsterdam: IOS Press; Berlin: Akademische Verlagsgesellschaft AKA GmbH, 2019, <<http://ebooks.iospress.nl/volume/mining-authoritativeness-in-art-historical-photo-archives-semantic-web-applications-for-connoisseurship>>.

30 Grafoo, <<https://essepuntato.it/graffoo>>.

Rappresentazione della conoscenza: creazione dei dati

Una volta completata l'attività di modellazione, è possibile tornare agli *item* per descriverli sulla base del modello concettuale progettato e implementato.

Può essere prodotta una tabella per ogni *item* (ad esempio un file CSV), in cui le colonne corrispondono alle triple soggetto/predicato/oggetto (linguaggio naturale per soggetti e oggetti; predicati sotto forma di proprietà come da ontologie scelte) e le righe sono gli *statement* o anche gli asserti.

- Creazione triple RDF. Primo passo per la creazione delle triple è il *namings*, ovvero l'identificazione univoca delle entità attraverso URI (anche se le istanze non sono referenziate).

La fase successiva è stabilire i prefissi per le ontologie scelte e scrivere le triple in RDF (in una delle possibili serializzazioni), cercando di collegare le entità e unire negli asserti le diverse lenti di osservazione: persona, luogo, data e soggetto/concetto.

È necessario anche includere collegamenti ad altre entità o concetti correlati, così come ipotizzati nel modello arricchito. Bisognerà prevedere collegamenti alle stesse autorità (owl:sameAs) per i nomi (ad esempio persone, luoghi, soggetti) e collegamenti ad altre risorse significative già presenti nel *LOD cloud*.

- Visualizzazione triple RDF. Ultima attività prevista è la visualizzazione del *knowledge graph* realizzato (ad esempio attraverso il tool RDF grapher)³¹ e la sua aggiunta alla documentazione come ultimo atto del *workflow*.

Un possibile scenario per il *workflow* LODLAM

Allo scopo di rendere ancora più chiare queste linee guida, viene di seguito accennato un possibile progetto nel settore LODLAM, assieme agli step di processo così come illustrati nella sezione precedente. Con Lease Morgan, ricordiamo che quando si realizza un progetto nel LODLAM:

moving forward with linked data is a lot like touring to Rome. There are many ways to get there, and there are many things to do once you arrive, but the result will undoubtedly improve your ability to participate in the discussion of the human condition on a world wide scale³².

Partiamo quindi da un'idea, definiamo lo scenario, scegliamo gli *item*, creiamo i primi modelli teorici a grafo, riconosciamo metadati e allineiamo i vocabolari, arricchiamo la nostra osservazione di dominio, elaboriamo il nostro modello concettuale riusando ontologie esistenti e, sulla base di questo modello, scriviamo adeguatamente le nostre triple RDF.

IDEA. Creazione dello scenario: la commedia di Dante (<https://it.wikipedia.org/wiki/Divina_Commedia>). Dettaglio (persone, luoghi, date, soggetti/concetti): i manoscritti fiorentini danteschi miniati del XIV-XV secolo conservati nelle biblioteche italiane.

31 RDF Grapher, <<http://www.ldf.fi/service/rdf-grapher>>.

32 Eric Lease Morgan; LiAM, *Linked archival metadata: a guidebook*, version 0.99. April 2014, <<http://infomotions.com/sandbox/liam/tmp/guidebook.pdf>>.

ITEM. Ricerca degli oggetti e delle relative descrizioni: una selezione di manoscritti (immagini in diversi repertori e descrizioni codicologiche), dipinti che rappresentano Dante, ritratti di miniatori e copisti, una mappa storica di Firenze, una lettera in cui si menziona il processo di produzione di un manoscritto, un libro moderno sulla miniatra fiorentina del XIV-XV secolo ecc.

MAPPA CONCETTUALE. Lavoro sui dati: il manoscritto “Riccardiano 1035 della Commedia” *conservato-in* “Biblioteca Riccardiana” *situata-in* Firenze *realizzato-nel* 1450 e *copiato-da* una persona chiamata “Giovanni Boccaccio” *raffigurato-in* “Giovanni Boccaccio, from the cycle Famous People” *realizzato-da* “Andrea del Castagno” ecc.

MODELLO E/R. Astrazione della vista sul dominio osservato: un Manoscritto (con i suoi attributi, ovvero i metadati) *copiato-da* una Persona (con i suoi attributi, ovvero i metadati) *raffigurata-in* una Opera d'arte (con i suoi attributi, ovvero i metadati) *conservata-in* un Museo (con i suoi attributi, ovvero i metadati) ecc.

ANALISI DEI METADATI. Studio degli standard: le istituzioni che hanno i manoscritti hanno usato TEI per la descrizione codicologica; OA è stato usato per descrivere il contenuto dei dipinti, DC per le categorie descrittive della mappa, EAD per la lettera, ISBD per la descrizione catalografica dell'*item* libro ecc.

ALLINEAMENTO. Mapping degli elementi: Persone (ad esempio nella forma di *agent*): DC:creator = EAD:creator = ICCD-F:author = ISBD:statement of responsibility = TEI:author, TEI:person, TEI:editor ecc.; Luoghi (ad esempio luoghi di creazione dell'*item*): DC:coverage = ISBD:place = TEI:placename; Date (le date andranno distinte sulla base del livello di osservazione del concetto di temporalità): DC:date; Soggetti (verificando se esiste un *value vocabulary* controllato): DC:subjects ecc.

MODELLO TEORICO. Arricchimento in linguaggio naturale: in relazione alle persone (chi: Dante, Boccaccio, Andrea del Castagno), potrebbe essere possibile concentrarsi, ad esempio, su dati prosopografici, ruoli, relazioni con oggetti di cui sono creatori, record di autorità, altre persone correlate ecc.; in merito ai luoghi (dove: Firenze), è possibile ad esempio identificarli su una mappa, geolocalizzarli, connetterli con le persone ecc.; circa le date (quando: 1450) è possibile ad esempio lavorare su altri eventi importanti accaduti nello stesso periodo, collegare le date a eventi su una linea temporale ecc.; sui concetti (cosa: copia, manoscritto, opere d'arte) è possibile ad esempio collegare una rete semantica lessicale esterna, trovare la forma controllata del nome, collegare parole chiave relative a un qualche livello d'osservazione (opera, espressione, manifestazione o unità), descrivere il processo creativo come un evento ecc.

MODELLO E/R ARRICCHITO. Arricchimento del grafo originario: è possibile ora aggiungere nuove entità e relazioni: una Persona *nata-in* un Luogo; un Manoscritto *prodotto-da* una Persona; un Evento *accaduto-in* una Data ecc.

MODELLO CONCETTUALE. Progettazione dell'ontologia: identificazione dei vocabolari pertinenti rispetto al modello teorico ideato, ad esempio: EAC-CPF per Person, CRM per Production, EDM per Event, PRO per Roles, PROV-O per Provenance, FRBR per Levels ecc.

REALIZZAZIONE RDF. Creazione degli URI, ad esempio da w3id.org:
<http://w3id.org/myproject/person/dante-alighieri>;
<http://w3id.org/myproject/event/manuscript-production/riccardiano1035>;
<http://w3id.org/myproject/time-span/1450>;
<http://w3id.org/myproject/object/portrait-boccaccio-del-castagno>

Scrittura delle triple, con @prefix and statements (nell'esempio che segue in Turtle) e collegamenti esterni:

```
@prefix crm: <http://www.cidoc-crm.org/cidoc-crm/> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .

<http://w3id.org/myproject/person/giovanni-boccaccio > a crm:E39_Actor ;
rdfs:label "Giovanni Boccaccio" ;
owl:sameAs <http://viaf.org/viaf/64002165> . # URI for 'Giovanni Boccaccio' from VIAF
```

Conclusioni

Già nel 2008 «a mashup culture built on shared resources was already very alive in the world of libraries, archives, and museums»³³. Negli ultimi venti anni, la discussione su LODLAM è diventata un argomento di ricerca interessante per la comunità legata alle scienze del libro e del documento. Il modello LOD come *workflow* tipico dei KOS, o anche come nuovo strumento di corredo nei sistemi di organizzazione della conoscenza, è una possibile soluzione all'analisi, al ragionamento critico, alla modellizzazione e all'arricchimento di risorse nel settore archivi, biblioteche e musei. Il ruolo delle DH aggiunge significato a questo processo, valorizzando il tema del lavoro a progetto su casi di studio reali in termini di ermeneutica, ovvero individuando nell'interpretazione del patrimonio culturale un possibile percorso di arricchimento semantico alla descrizione delle fonti.

Il corso di KO and CH nel corso di laurea DHDK è un possibile passo nella direzione di un rafforzamento del dialogo tra il mondo della Library and information science e il dominio delle DH.

RIFERIMENTI BIBLIOGRAFICI

[1] Brier Stephen, *Where's the pedagogy? The role of teaching and learning in the digital humanities*. In: *Debates in the digital humanities*, new edition, edited by Matthew K. Gold. Minneapolis: University of Minnesota Press, 2012, p. 390-412, <<https://www.jstor.org/stable/10.5749/j.ctttv8hq.25>>, DOI: 10.5749/j.ctttv8hq.25.

33 Jon Voss, *Radically open cultural heritage data on the web*. In: *Museums and the web 2012* (San Diego (CA), USA, 11th-14th April 2012). 25th March 2012, <http://www.museumsandtheweb.com/mw2012/papers/radically_open_cultural_heritage_data_on_the_w>.

[2] *A companion to digital humanities*, edited by Susan Schreibman, Ray Siemens, John Unsworth. Oxford: Blackwell, 2004, <<http://www.digitalhumanities.org/companion>>.

[3] *Computing in humanities education: a European perspective*, edited by Koenraad De Smedt [et al.]. Bergen: University of Bergen. HIT-centre, 1999, <<http://korpus.uib.no/humfak/AcoHum/book>>.

[4] Daquino Marilena, *Mining authoritativeness in art historical photo archives: semantic web applications for connoisseurship*. Amsterdam: IOS Press; Berlin: Akademische Verlagsgesellschaft AKA GmbH, 2019, <<http://ebooks.iospress.nl/volume/mining-authoritativeness-in-art-historical-photo-archives-semantic-web-applications-for-connoisseurship>>, DOI: 10.3233/SSW190006-mono.

[5] Daquino Marilena [et al.], *Enhancing semantic expressivity in the cultural heritage domain: exposing the Zeri Photo Archive as linked open data*, «Journal on computing and cultural heritage», 10 (2017), n. 4, article 21, <<https://dl.acm.org/doi/10.1145/3051487>>, DOI: 10.1145/3051487.

[6] *Digital humanities pedagogy: practices, principles and politics*, edited by Brett D. Hirsch. Cambridge: Open Book Publishers, 2012, <<https://www.openbookpublishers.com/product/161>>, DOI: 10.11647/OBP.0024.

[7] Gonano Ciro Mattia [et al.], *Zeri e LODE: extracting the Zeri photo archive to linked open data: formalizing the conceptual model*. In: *2014 IEEE/ACM Joint Conference on Digital Libraries* (London, United Kingdom, 8th-12th September 2014). [Piscataway]: Institute of Electrical and Electronical Engineers, 2014, p. 289-298, <<https://ieeexplore.ieee.org/document/6970182>>, DOI: 10.1109/JCDL.2014.6970182.

[8] Heath Tom; Bizer Christian, *Linked data: evolving the web into a global data space*. [San Rafael (CA)]: Morgan & Claypool, 2011, <<https://www.morganclaypool.com/doi/abs/10.2200/S00334ED1V01Y201102WBE001>>, DOI: 10.2200/S00334ED1V01Y201102WBE001.

[9] Hjørland Birger, *What is knowledge organization (KO)?*, «Knowledge organization», 35 (2008), n. 2/3, p. 86-101, <<https://www.nomos-elibrary.de/10.5771/0943-7444-2008-2-3-86/what-is-knowledge-organization-ko-volume-35-2008-issue-2-3?page=1>>, DOI: 10.5771/0943-7444-2008-2-3-86.

[10] Hyvönen Eero, *Publishing and using cultural heritage linked data on the semantic web*. [San Rafael (CA)]: Morgan & Claypool, 2012, <<https://www.morganclaypool.com/doi/abs/10.2200/S00452ED1V01Y201210WBE003>>, DOI: 10.2200/S00452ED1V01Y201210WBE003.

[11] Isaac Antoine [et al.], *Library linked data incubator group: datasets, value vocabularies, and metadata element sets: W3C Incubator Group report*. October 2011, <<https://www.w3.org/2005/Incubator/ld/XGR-ld-vocabdataset-2011025/>>.

[12] Morgan Eric Lease; LiAM, *Linked archival metadata: a guidebook*, version 0.99. April 2014, <<http://infomotions.com/sandbox/liam/tmp/guidebook.pdf>>.

[13] Riley Jenn, *Seeing standards: a visualization of the metadata universe*. 2009-2010, <<http://jennriley.com/metadatamap/>>.

[14] Rockwell Geoffrey; Sinclair Stefan, *Acculturation and the digital humanities community*. In: *Digital humanities pedagogy: practices, principles, politics*, edited by Brett D. Hirsch. Cambridge: Open Book Publishers, 2012, p. 177-211, <<https://www.openbookpublishers.com/product/161>>, DOI: 10.11647/OBP.0024.

[15] Taylor Arlene G., *The organization of information*. Englewood (CO): Libraries Unlimited, 1999.

[16] *Teaching with digital humanities: tools and methods for Nineteenth-Century American literature*, edited by Jennifer Travis, Jessica DeSpain. Urbana: University of Illinois Press, 2018, <<https://www.jstor.org/stable/10.5406/j.ctv8bt13m>>, DOI: 10.5406/j.ctv8bt13m.

[17] Thaller Manfred; Sahle Patrick, *Digitale Geisteswissenschaften*. Cologne: Cologne Center for eHumanities, 2011, <<http://www.cceh.uni-koeln.de/Dokumente/BroschuereWeb.pdf>>.

[18] Tomasi Francesca, *Digital humanities and digital knowledge (DHDK): international second cycle/master degree*, «Umanistica digitale», 2 (2018), p. 225-231, <<https://umanisticadigitale.unibo.it/article/view/7862>>, DOI: 10.6092/issn.2532-8816/7862.

[19] Voss Jon, *Radically open cultural heritage data on the web*. In: *Museums and the web 2012* (San Diego (CA), USA, 11th-14th April 2012). 25th March 2012, <http://www.museumsandtheweb.com/mw2012/papers/radically_open_cultural_heritage_data_on_the_w>.

[20] Zorich Diane M.; Waibel Gunter; Erway Ricky, *Beyond the silos of the LAMs: collaboration among libraries, archives and museums*. Dublin (OH): OCLC Research, September 2008, <<http://www.oclc.org/content/dam/research/publications/library/2008/2008-05.pdf>>.

Articolo proposto il 28 maggio 2020 e accettato il 9 luglio 2020.

ABSTRACT AIB studi, 60 n. 2 (maggio/agosto 2020), p. 411-425. DOI 10.2426/aibstudi-12068
ISSN: 2280-9112, E-ISSN: 2239-6152

FRANCESCA TOMASI, Alma mater studiorum Università di Bologna, Dipartimento di filologia classica e italianistica, e-mail francesca.tomasi@unibo.it.

Digital humanities e organizzazione della conoscenza: una pratica di insegnamento nel LODLAM

La mancanza di una definizione unanimemente condivisa su cosa significhi insegnare *digital humanities* (DH) ci permette di riflettere su obiettivi e metodologie formative, con l'obiettivo di ragionare sull'identità dell'umanista digitale. Il corso di laurea magistrale internazionale in Digital humanities and digital knowledge (DHDK), attivo presso l'Università di Bologna, ha lo scopo di contribuire alla discussione, proponendo un possibile modello di insegnamento delle DH.

Data l'esperienza di questo progetto didattico, il corso di Knowledge organization and cultural heritage (KO and CH), una delle attività obbligatorie del primo anno del corso di laurea, è stato progettato avendo in mente un principio: enfatizzare il ruolo delle teorie, metodologie e tecniche delle DH nell'insegnare come affrontare quel percorso che va dalla creazione dei dati alla rappresentazione della conoscenza, e gestire il processo come un flusso di lavoro iterativo in ottica progettuale.

Questo articolo mira a introdurre lo scopo e il contenuto del corso di KO and CH. In particolare, si vuole focalizzare sulle linee guida richieste per creare un progetto nel contesto dei linked open data, lavorando

su dati, modelli e metodi provenienti dal dominio di biblioteche, archivi e musei (LODLAM). I principi che governano il percorso necessario a produrre LOD sono affrontati nei termini della riflessione sulla metodologia per la gestione dei sistemi di KO nel contesto del LAM. L'uso di teorie e metodologie delle DH, a livello di interpretazione delle memorie del patrimonio culturale, è il valore aggiunto alla progettazione, modellizzazione e implementazione del lavoro a progetto.

A digital humanities methodology for teaching knowledge organization and LAM

The lack of a definition of what it means to teach Digital Humanities (DH) that is unanimously agreed upon allows us to reflect on educational objectives and methodologies with the aim of shaping the identity of the digital humanist. The opening of the international Master's degree program in Digital Humanities and Digital Knowledge (DHDK) at the University of Bologna aims to contribute to the discussion by developing a possible DH teaching model.

Setting out from the experience developed through this project, the Knowledge Organization and Cultural Heritage (KO and CH) course, one of the first year's compulsory modules, has been designed following a specific principle. While teaching how to move from data to knowledge and how to manage this process as an iterative workflow, this course emphasises the role of DH theories, methodologies and techniques.

This paper aims to introduce the scope and content of the KO and CH course. In particular, it focuses on the guidelines required to create a project in the domain of Linked open data, by working on data, models and methods coming from libraries, archives and museums (LODLAM). Linked open data (LOD) is considered to be the primary methodology for managing KO issues in the context of LAM. For what concerns the interpretation of cultural heritage memories, the implementation of DH theories and methodologies is the added value to the project's design, modelling and implementation.